

# Curve Fitting Tool

Grigore Stamatescu  
*grigore.stamatescu@upb.ro*

## Introducere

Curve Fitting Tool este utilitarul MATLAB care oferă interfețe grafice și funcții pentru linia de comandă pentru atribuirea de funcții și suprafețe seriilor de date. Utilitarul permite realizarea de analize de date exploratorii, pre-procesarea și post-procesarea datelor, compararea modelelor candidate și înlăturarea excepțiilor. Pot fi implementate regresii folosind librăriile existente de modele liniare și neliniare sau pot fi specificate propriile ecuații. Librăria oferă parametri de rezolvare optimizați și condiții inițiale pentru îmbunătățirea calității fiturilor. Pachetul suportă și tehnici de modelare neparametrice, cum sunt spline, interpolarea și netezirea.

Caracteristici cheie:

- Instrumente grafice pentru fitarea curbelor și suprafețelor
- Regresii liniare și neliniare cu ecuații personalizate
- Bibliotecă de modele de regresii cu puncte inițiale optimizate și parametri de rezolvare
- Metode de interpolare, inclusiv B-splines, splines tensor-produs, etc.
- Tehnici de netezire, inclusiv splines de netezire, regresii localizate, filtre Savitsky-Golay și medii alunecătoare
- Rutine de pre-procesare, inclusiv înlăturarea extremelor și secționare, scalare și ponderarea datelor
- Rutine de post-procesare, inclusiv interpolare, extrapolare, intervale de încredere, integrale și derivate

## Utilizare

Pentru deschiderea Curve Fitting Tool, introduceți următoarea comandă în linia de comenzi MATLAB:

```
» cftool
```

iar utilitarul se va deschide sub forma unei ferestre noi (vezi Figura 1).

Fereastra principală are 5 butoane:

**Data...:** Deschide o fereastră nouă care permite utilizatorului importarea de date din spațiul de lucru MATLAB în mediul Curve Fitting. Acesta va fi denumit set de date.

**Fitting...:** Această fereastră permite alegerea funcției de fitare și a parametrilor ei.

**Exclude...:** Cu această interfață putem exclude anumite puncte de date din setul de date.

**Plotting...:** Aici sunt selectate opțiunile de afișare pentru seturile de date și graficele generate. Pot fi alese selectiv cele pe care dorim să le reprezentăm grafic.

**Analysis...:** Interfață grafică pentru evaluate funcției atribuite/fitate setului de date.

### 1. Data...

La apăsarea butonului 'Data...', se deschide o nouă fereastră (Figura 2) și putem alege x, y și variabilele pondere, numele setului de date și acesta poate fi creat prin apăsarea butonului 'Create data set'. Setul de date va fi afișat în listă și va putea fi utilizat în pașii următori.

Acest scurt exemplu nu include setare unor valori pentru ponderi deoarece variabilele x și y su fost create. În cazul analizei unor date, cum este măsurarea la un interval de timp finit e.g. la fiecare secundă, parametrul

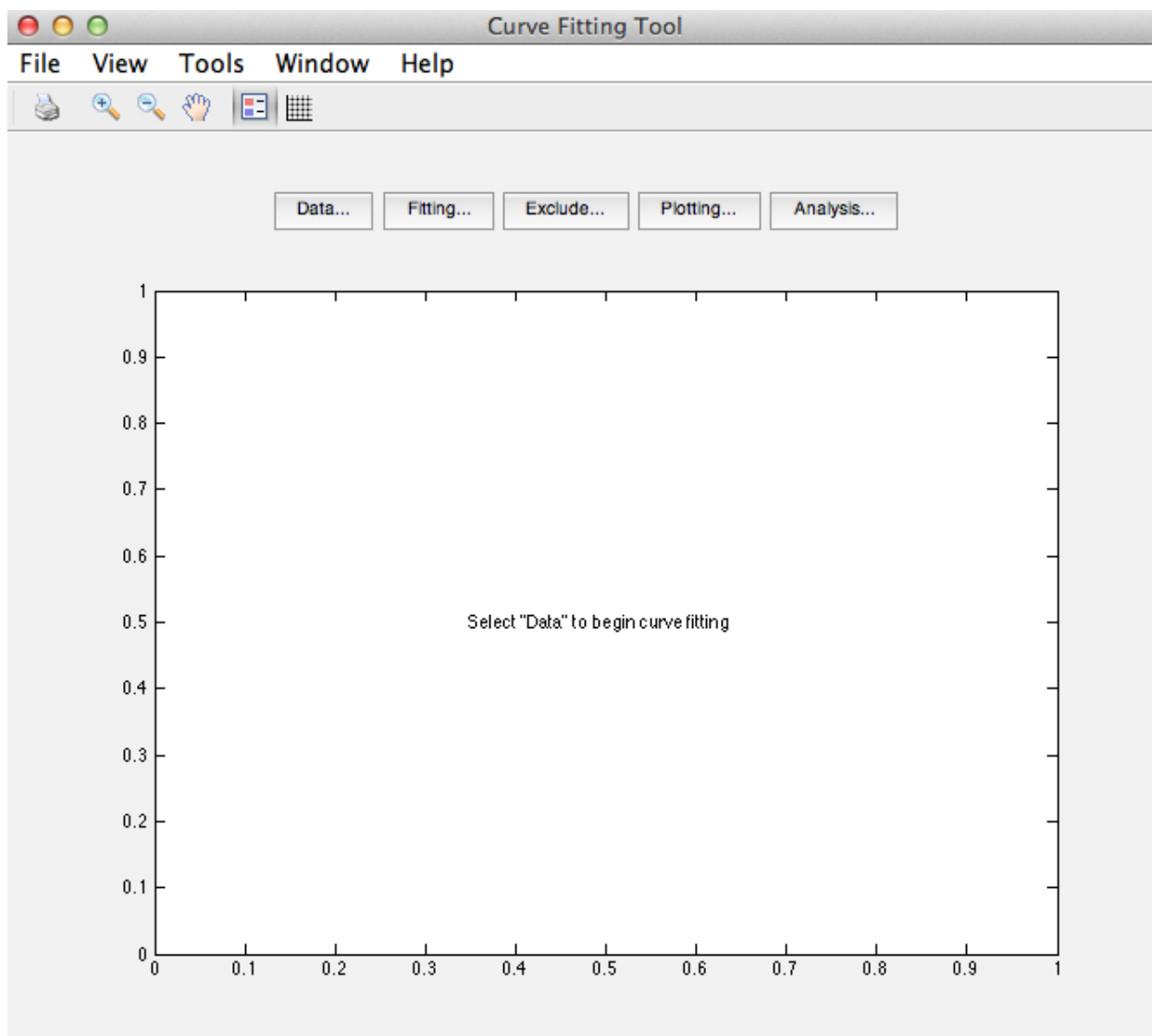


Figure 1: Fereastra principală - Curve Fitting Tool

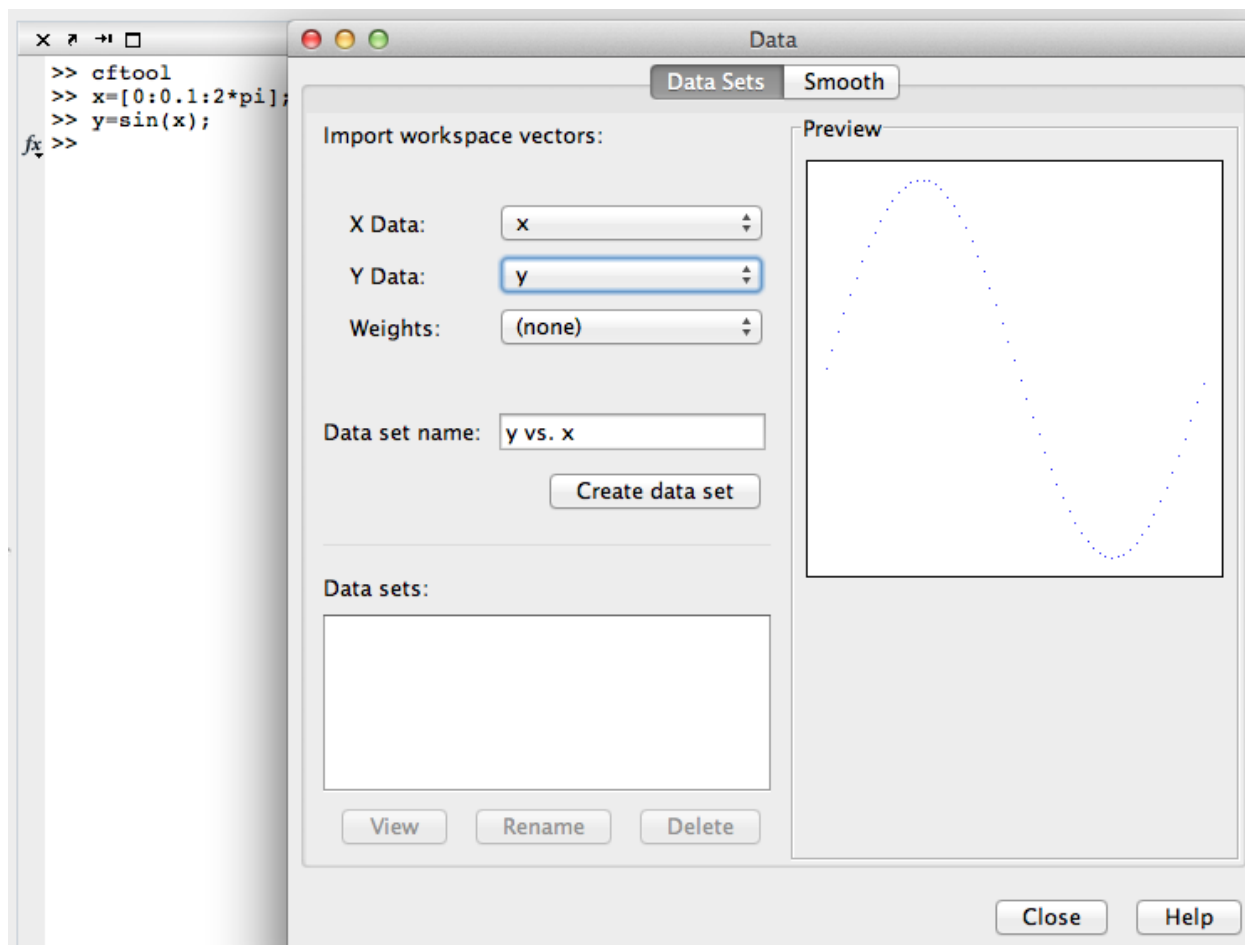


Figure 2: Crearea unui nou set de date

$x$  este unul ales (parametrul liber). Parametrul  $y$  este unul măsurat (dependent) care prezintă erori diferite între punctele de date. Fitarea normală, fără ponderi, consideră toate punctele egale dar în cazul în care unele puncte au bare de erori mai mari, relativ la punctele celelalte, asta înseamnă că sunt mai puțin precise și că fitarea ar trebui să le ia în considerare mai puțin în calcule. Acesta este rolul ponderilor. Punctele de date mai precise primesc ponderi mai mari în raport cu punctele de date mai puțin precise.

Modul de a crea ponderile, corelat cu logica expusă mai sus, este ca ponderea unui punct de date să fie inversul pătratului barei de eroare. Exprimată matematic: dacă  $w_i$  este ponderea și  $\sigma_i$  este eroarea punctului de date  $i$ , atunci:

$$w_i = \frac{1}{\sigma_i^2} \quad (1)$$

## 2. Fitting...

Pentru a fi datele, trebuie mai întâi deschis editorul Fitting prin apăsarea butonului corespunzător. Se alege 'New fit...', numele fitului (cel standard fiind compus din fit și un număr), se alege setul de date predefinit, funcția de fitare și se apasă 'Apply'.

Există o serie de funcții de fitare incluse, care sunt grupate după tip: "Sumă de sinusoidă", "Polinomiale", etc. Din fiecare grup trebuie aleasă funcția specifică care se dorește a fi folosită pentru fitare. În acest caz vom alege o funcție din grupul "Sumă de sinusoidă".

La apăsarea butonului 'Apply', este calculat fitul și rezultatele sunt afișate în zona dedicată a editorului și reprezentare grafic în fereastra principală a Curve Fitting Tool. Figura 3 prezintă utilizarea Fit Editor și rezultatul operațiilor efectuate. Inspecția vizuală a rezultatului arată că fitarea este bună și coeficientul de determinare R-square și R-square ajustat confirmă acest lucru. Putem vedea de asemenea că amplitudinea,  $a1$ , este 1, frecvența (sau mai exact,  $\omega = 2\pi f$ , care este  $b1$  dacă variabila  $x$  este timpul) este, de asemenea, 1 și faza,  $c1$ , este 0, identic cu parametrii aleși la crearea vectorului  $y$ ,  $y = \sin(x)$ .

Numerele din paranteză, de lângă valorile parametrilor (coeficienții) sunt intervalele de încredere - eroarea fitului. Deoarece fitul nostru se potrivește exact setului de date, aceste valori sunt zero. Această observație nu va mai fi valabilă în exemplele ce vor urma.

La modul general, algoritmul de fitare pornește de la niște valori ale constantelor funcției de fitare,  $a1$ ,  $b1$  și  $c1$  în acest caz, și încearcă să convergă de la acești parametri inițiali, la cei finali. Uneori parametrii inițiali nu sunt destul de apropiați de cei finali pentru a converge la aceștia sau nu converg deloc. Meniul 'Fit options...' permite modificarea acestor parametri.

În cazul în care o funcție de fitare de care avem nevoie nu există în lista de funcții, putem alege ecuații personalizate prin butonul 'Custom equations' și definirea funcției dorite.

'Save to workspace...' va salva datele de fitare în spațiul de lucru și de acolo vom putea lucra cu aceste date în linia de comandă.

## 3. Exclude..., Plotting... și Analysis...

Primele două opțiuni vor deschide ferestre noi care permit eliminarea unor puncte de date din fit (ele există în continuare în setul de date) sau eliminarea unor seturi de date sau fituri din reprezentarea grafică.

În meniul de analiză putem obține valoarea funcției fitate în anumite puncte alese și intervalele de încredere. Pot fi determinate de asemenea primele două derivate și integrată funcția. Aceste funcționalități vor fi detaliate în secțiunea de exemple.

## 4. Reziduuri

Accesarea meniului 'View' → Residuals → 'Scatter Plot' va deschide un nou grafic care va prezenta reziduurile, diferența finală dintre punctele de date inițiale și fit. Figura 4 ilustrează rezultatele pentru exemplul dat.

La analiza reziduurilor este important să vedem că acestea sunt împrăștiate aleatoriu și uniform în jurul

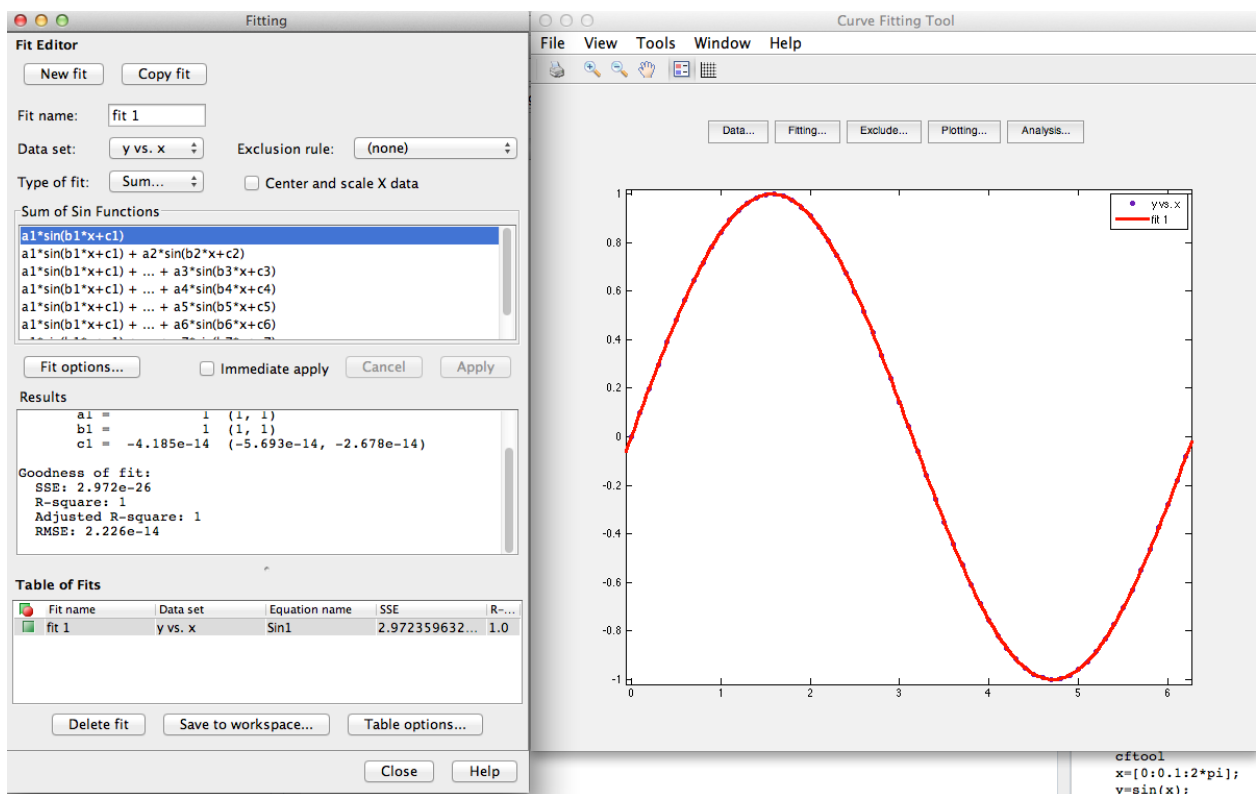


Figure 3: Fitarea datelor cu o funcție  $y = a_1 \cdot \sin(b_1 \cdot x + c_1)$

valorii de zero. Asta presupune ca să nu fie pozitive într-o parte a fitului și negative în altă parte, sau ca valoarea lor nu este mică sau mare în zone diferite.

În graficul prezentat, aceste criterii nu există, dar dacă analizăm valoarea reziduurilor, de până la  $5 \cdot 10^{-14}$ , putem înțelege că aceasta reprezintă doar o eroare de aproximare la calculul funcției *sin* deoarece calculele efectuate de calculator nu au precizie infinită.

### Exemple

1. Fit liniar simplu, cu bare de eroare

În acest exemplu vor fi fitate 5 puncte de date cu o funcție liniară. Introduceți următoarele comenzi:

```
>>x=[1 2 3 4 5];  
>>y=[1 2 3 4 4.5];
```

și barele de erori și ponderile sunt:

```
>> err=[0.1 0.2 0.3 0.4 0.5];  
>> weight=1./err.^2;
```

Se poate observa cum este generată o linie dreaptă, cu panta 1 în care ultimul punct pare a fi greșit. O opțiune este excluderea acestui punct din grafic, dar nu este cea preferată. Vom arăta cum adăugarea ponderilor are un efect similar.

În Figura 5 sunt ilustrate două fituri, cu (albastru) și fără (roșu) ponderi. În editor putem vedea rezultatul calculelor pentru setul de date cu ponderi, denumit fit 2, care arată parametrii de fitare 1 și 0, după cum era de așteptat. Atunci când fitul nu este perfect, limita de încredere are valori diferite. Între cele două ferestre sunt listate și comenzile MATLAB folosite, important fiind în special modul de definire al ponderilor. Erorile trebuie calculate separat, reprezentând în acest caz 10% din valoarea lui x.

În continuare, deschideți fereastra de analiză. Alegeți “fit 2 y vs. x with weights” și setați punctele de analiză la 0:0.5:10, între 0 și 10, cu incrementul 0.5. Ulterior alegeți “Evaluate fit at Xi” și “For function”. Vom adăuga și “Plot results” și “Plot data set” și apăsăm “Apply”. Rezultatul acestor operații este prezentat în Figura 6.

Se poate vedea cum fitul este relativ precis acolo unde avem puncte de date, dar pe domeniul fără puncte de date acesta începe să fie mai puțin precis. Urmând același mod de lucru pentru setul de date fără ponderi, vom observa o deteriorare a predicției.

În continuare va fi realizat un grafic care include și barele de eroare alături de funcția de fitare și limitele de încredere.

Salvați datele de analiză în spațiul de lucru cu numele standard “analysisresults1” și comanda “Save to Workspace”. Faceți dublu clic pe această variabilă în spațiul de lucru pentru a afișa conținutul ei.

Introduceți următoarele comenzi în linia de comandă MATLAB:

```
>> figure % deschide o nouă fereastră pentru grafice  
>> errorbar(x,y,err, 'ro') % reprezintă datele sub formă de cercuri, alături de barele de eroare cu roșu  
>> hold on % Reprezintă grafic următoarele rezultate, fără a le șterge pe cele vechi  
>> plot(analysisresults1.xi, analysisresults1.yfit, 'b-', analysisresults1.xi, analysisresults1.lower, 'b:', analysisresults1.xi, analysisresults1.upper, 'b:') % reprezintă fitul cu o linie albastră continuă și limitele de încredere cu o linie albastră întreruptă  
>> hold off % eliberează figura
```

Figura 7 prezintă rezultatele obținute. Aceasta conține și elemente vizuale suplimentare cum sunt titluri pe grafic și pe axe, legenda și o căsuță de text în care a fost introdus rezultatul calculelor aferente fitării. Elementele pot fi adăugate utilizând instrumentele grafice “Plot Tools” din figura MATLAB.

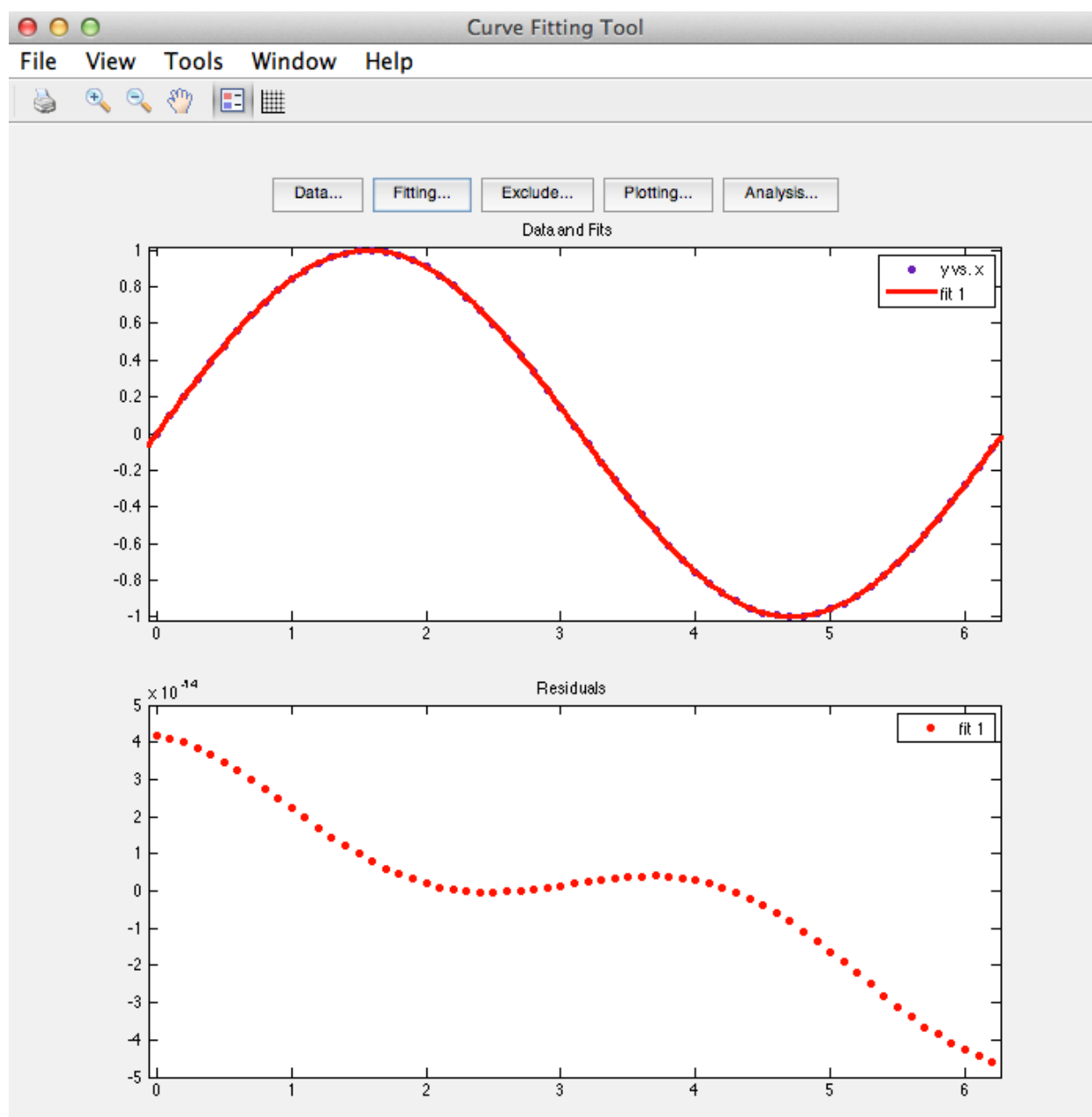


Figure 4: Graficul cu afișarea reziduurilor

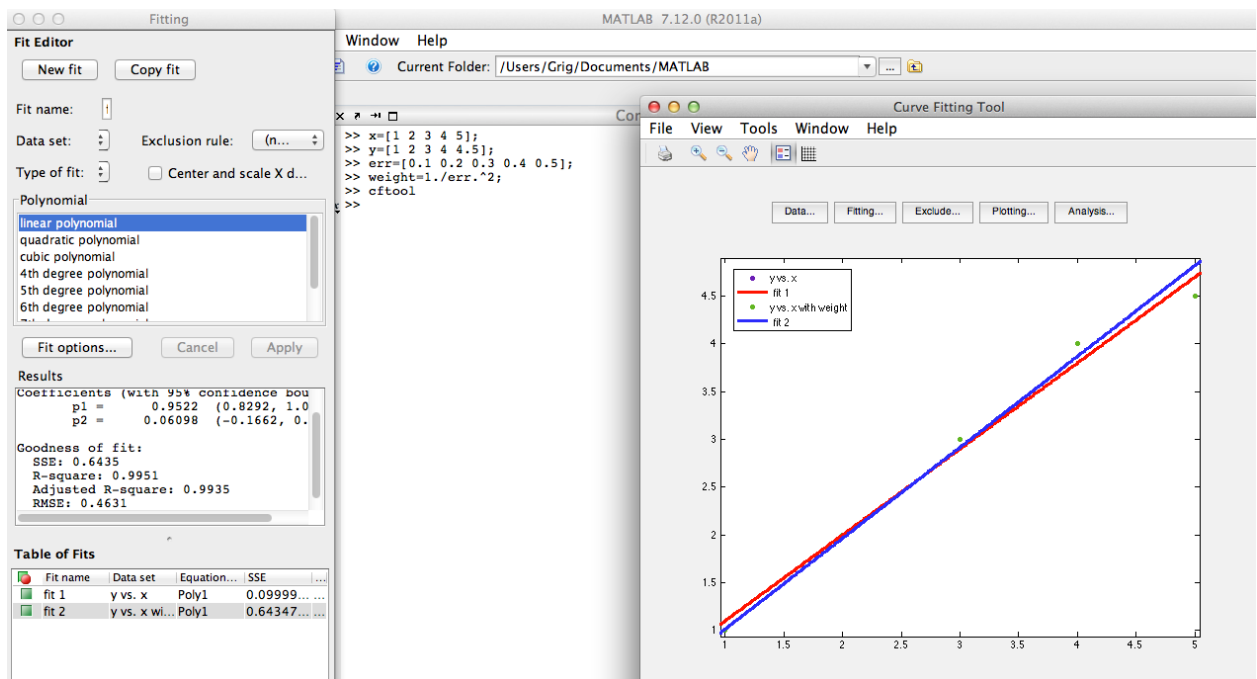


Figure 5: Fitare cu funcție liniară: cu (albastru) și fără (roșu) ponderi.

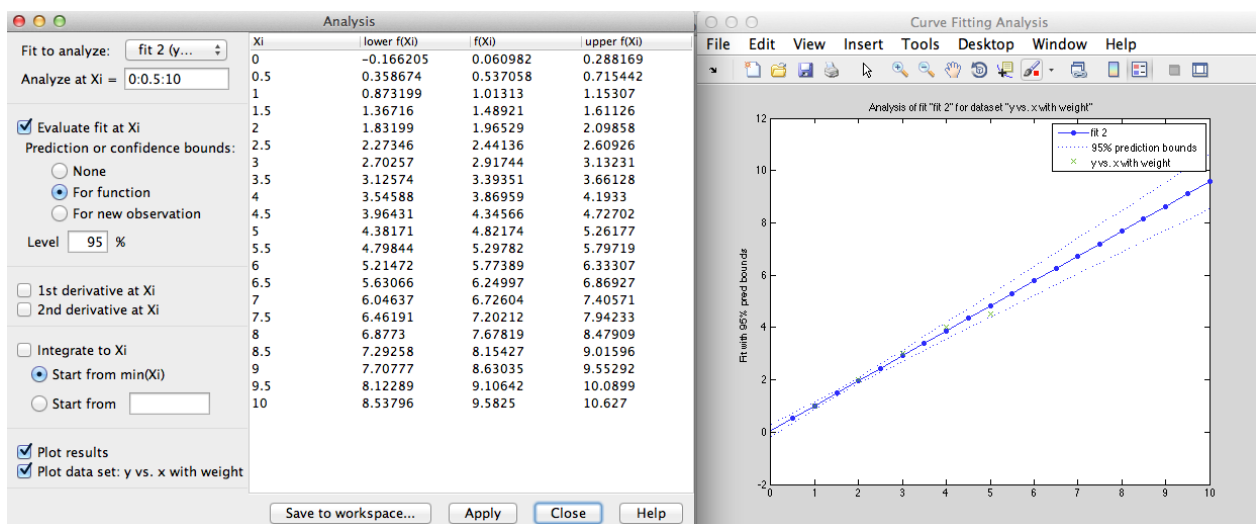


Figure 6: Fereastra de analiză și rezultatele obținute



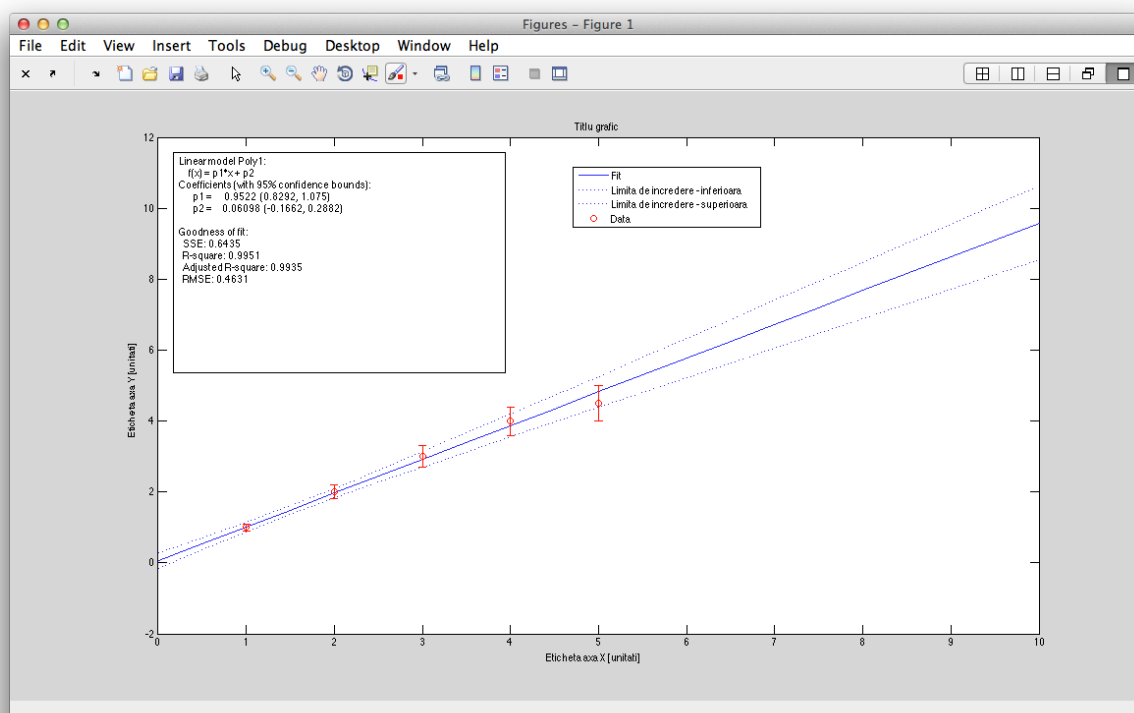


Figure 7: Grafic cu setul de date, bare de eroare, funcția de fitare și limite de încredere

## 2. Fit avansat

Acest exemplu prezintă fitarea unui set de date cu o funcție de tipul  $\sin^2$ . Funcția de fitare nu este inclusă în opțiunile standard și va trebui creată. De asemenea, vor trebui definiți alți parametri inițiali. În acest exemplu este folosit un set de date afectate de zgomot. Acesta este introdus în linia de comandă MATLAB, după cum urmează:

```

>>x=[0:0.1:2*pi];
>>y=8*sin(2*pi*0.2*x+1).^2+randn(size(x));
    
```

Figura prezintă setul de comenzi MATLAB folosit pentru crearea datelor și adăugarea unor valori aleatorii. Datele sunt un  $\sin^2$  cu amplitudinea 8, frecvența 0.2 și faza 1 iar scopul este obținerea acestor valori din funcția de fitare.

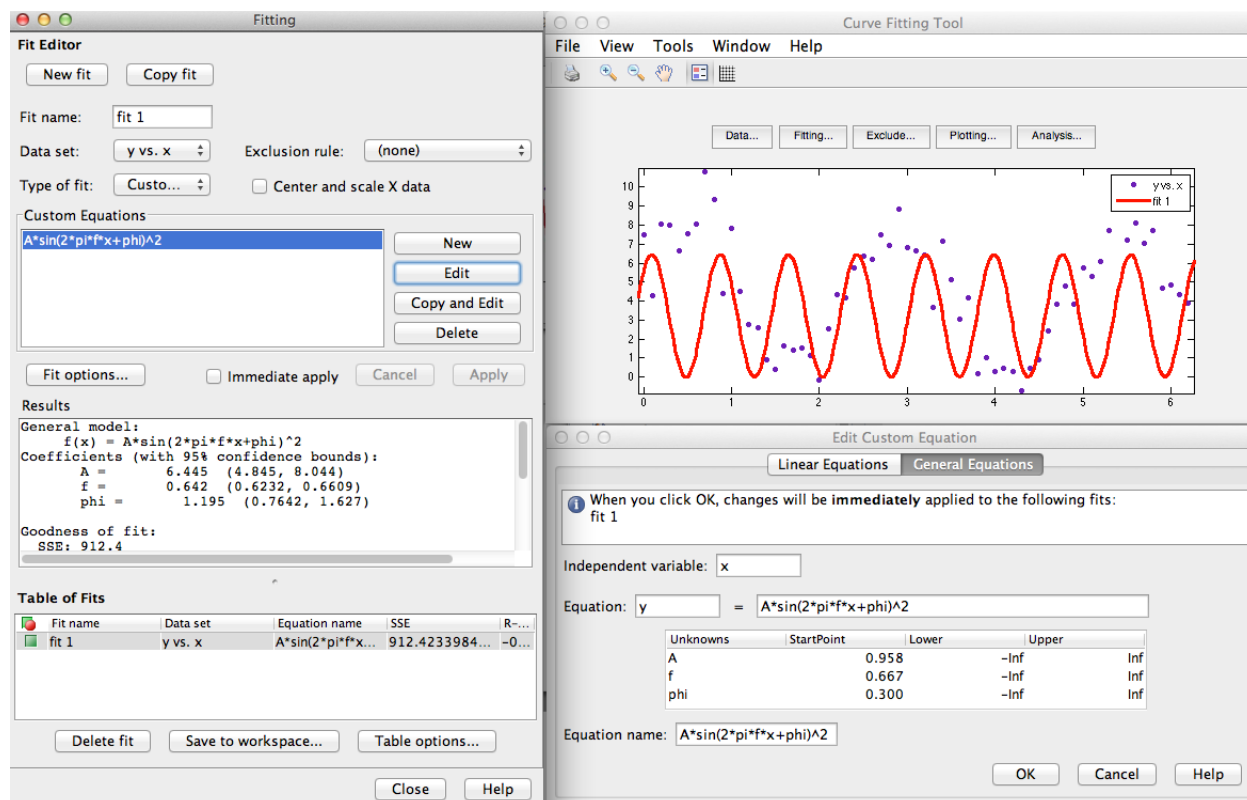


Figure 8: Fitare cu funcție tip  $\sin^2$

Pentru a ajunge la rezultatele ilustrate, trebuie create variabilele în spațiul de lucru MATLAB, deschis utilitarul Curve Fitting Tool și creat setul de date. Ulterior, va fi creată fitarea folosind “New Fit” și o ecuație introdusă de la tastatură ‘Custom Equation’ → New Equation → ‘Create Custom Equation’ → ‘General Equations’. Aici trebuie inserați și parametri inițiali de fitare, pe care îi vom lăsa deocamdată nemodificați. După apăsarea OK, ecuația va fi adăugată la lista de ecuații personalizate din editor. Aplicare modificărilor va conduce la rezultatele reprezentate mai sus. Acestea însă nu sunt corecte, după cum este indicat și în zona de rezultate a fitului.

Prin apăsarea “Fit Options...” se va deschide o nouă fereastră unde pot fi modificați parametrii de fitare. După identificarea corectă a acestora, afișați și reziduurile pentru a verifica distribuția lor aleatoare.

## Referințe

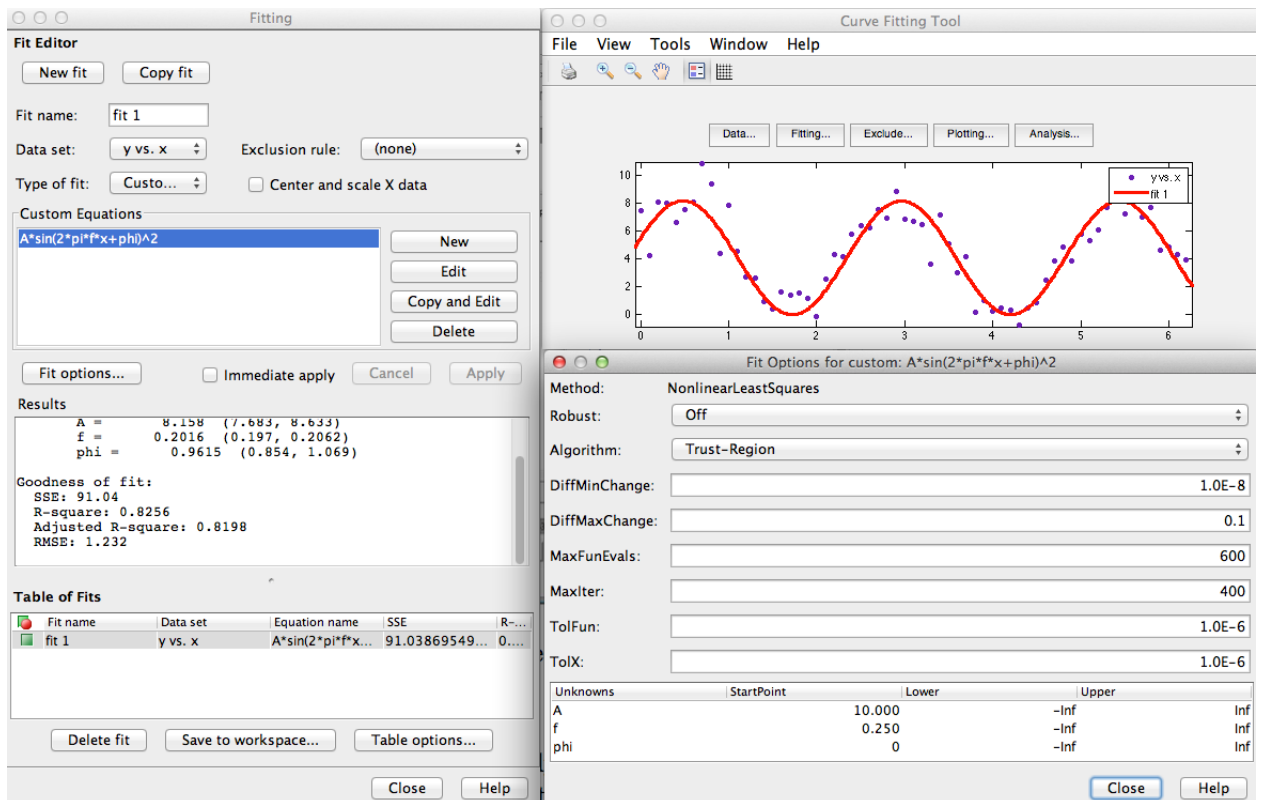


Figure 9: Modificarea parametrilor de fitare

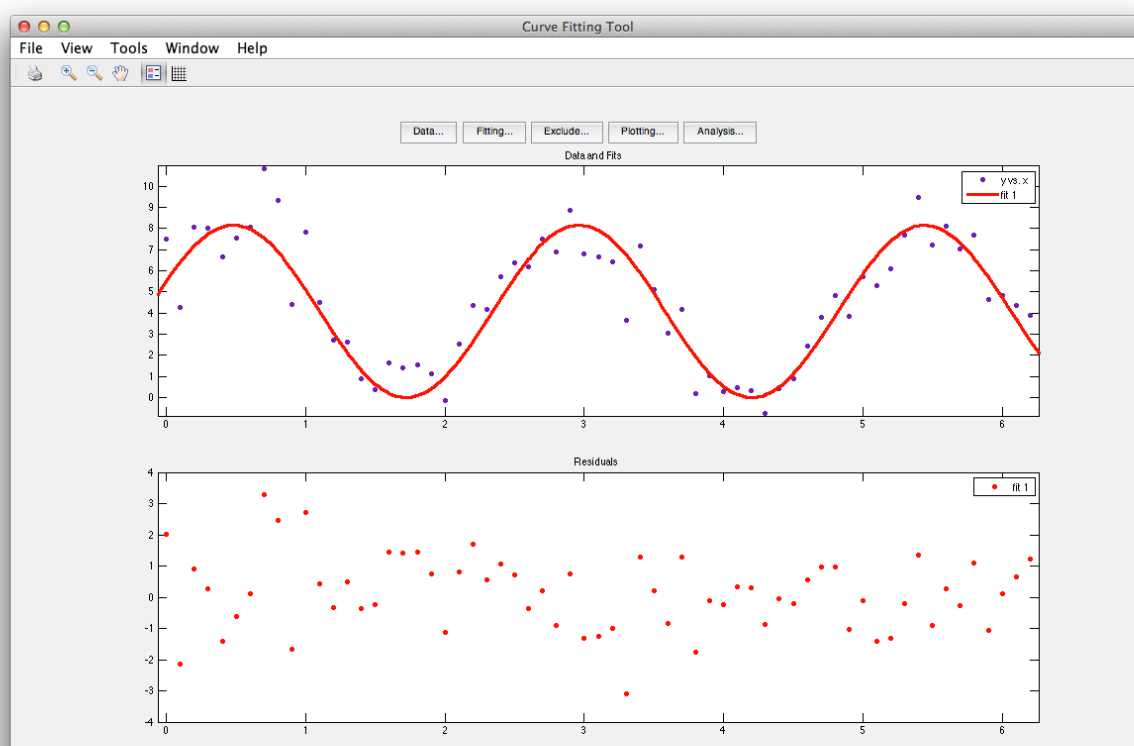


Figure 10: Rezultate finale cu erori reziduale

[1] MathWorks, *MATLAB Curve Fitting Toolbox User's Guide*, R2011a.